

Systematic design of a learning environment for domain-specific and domain-general critical thinking skills

Dawit Tibebe Tiruneh¹ · Ataklti G. Weldelessie² ·
Abraham Kassa³ · Zinaye Tefera³ · Mieke De Cock⁴ ·
Jan Elen¹

Published online: 31 December 2015

© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract Identifying effective instructional approaches that stimulate students' critical thinking (CT) has been the focus of a large body of empirical research. However, there is little agreement on the instructional principles and procedures that are theoretically sound and empirically valid to developing both domain-specific and domain-general CT skills. The purpose of this study was to examine the effectiveness of systematically designed subject matter instruction in stimulating the development of domain-specific and domain-general CT skills, and to investigate the relationship between the two. The study employed a pretest–posttest quasi-experimental design with two conditions: 45 students participated in an experimental condition and 44 students in a control condition. A learning environment, in the context of a freshman physics course, was designed according to the First Principles of Instruction model. The experimental condition followed the designed learning environment, while the control condition followed regular subject matter instruction that was not designed according to the First Principles of Instruction model. The experimental condition scored significantly higher than the control condition on a domain-specific CT test. The results also showed that better performance on a domain-specific CT test explained a significant proportion of the variance on a domain-general CT test. However, the experimental learning environment did not result in a significantly greater pretest–posttest improvement in the acquisition of domain-general CT skills compared to the control learning environment. Instructional design principles that may contribute to the

✉ Dawit Tibebe Tiruneh
dawittibebe.tiruneh@ppw.kuleuven.be

¹ Centre for Instructional Psychology and Technology, KU Leuven, Dekenstraat 2, Box 3773, 3000 Leuven, Belgium

² Science, Engineering and Technology Group, KU Leuven Group T, Vesaliusstraat 13, Box 2600, 3000 Leuven, Belgium

³ Department of Physics, College of Science, Bahir Dar University, Box 79, Bahir Dar, Ethiopia

⁴ Department of Physics and Astronomy and LESEC, KU Leuven, Celestijnenlaan 200C, Box 2406, 3001 Leuven, Belgium

present understanding of the integration of CT skills within the regular subject matter instruction are discussed.

Keywords Critical thinking · Domain-specific and domain-general critical thinking · Instructional design · Physics · Higher education

Introduction

As a vital twenty-first century skill, development of critical thinking (CT) is one of the major goals of higher education (Association of American Colleges and Universities 2005; Halpern 2014; van Gelder 2005). CT involves the ability to clearly and precisely raise vital questions, gather relevant information and reach well-reasoned conclusions, make accurate decisions, assess the credibility of sources, identify cause–effect relationships, and effectively communicate with others in figuring out solutions (Ennis 1989; Halpern 1998). Proficiency in CT is linked with success in college (Williams et al. 2004; Zohar and Nemet 2002), improved decision-making with regard to complex, real-life problems (Dwyer et al. 2012; Facione 1990a), and more generally with a tendency to become a more active and informed citizen (Butler et al. 2012; Halpern 2014).

Although widely recognized as desirable, CT competence is often found to be low among students in higher education (e.g., Arum and Roksa 2011; Billing 2007; Pascarella and Terenzini 2005; van Gelder 2005). Identifying effective ways to develop students' CT has therefore been the focus of a large body of intervention-based research. Most previous efforts to address the challenge of CT development took place in a context in which general CT skills were taught separately from regular subject matter domains (for reviews, see McMillan 1987; Pascarella and Terenzini 2005). In this approach, the ability to think critically is seen as independent of the acquisition of knowledge and skills of a particular subject matter domain. However, this point of view has become less dominant in recent years. Empirical attempts to develop students' CT have shifted mainly towards embedding CT skills within subject matter domains (for reviews, see Abrami et al. 2008, 2015; Niu et al. 2013; Tiruneh et al. 2014).

The accompanying expectation has been that embedding CT skills within a subject matter domain will enable the acquisition of CT skills that are applicable to a wide variety of thinking tasks within the domain in question and that it will facilitate transfer to other problems in everyday life (Perkins and Salomon 1989; Resnick et al. 2010; van Merriënboer 1997). Successful teaching of CT within a subject matter domain is in other words expected to result in the development of both domain-specific and domain-general CT skills. Domain-specific CT refers to the ability to think critically in a domain that requires specific subject matter expertise (e.g., McPeck 1990), while domain-general CT refers to the ability to think critically in a domain that requires knowledge of everyday life (e.g., Ennis 1989).

Although extensive empirical studies have focused on developing domain-general CT skills (see Abrami et al. 2008), research into the acquisition of domain-specific CT skills has largely been lacking (Fischer et al. 2014; Pascarella and Terenzini 2005; Tiruneh et al. 2014). Aside from a few successful attempts in the domain of psychology (e.g., Penningroth et al. 2007; Williams et al. 2004), there is a dearth of empirical evidence with respect to the question of whether embedding CT skills within a subject matter domain

promotes the development of CT in specific domains of science and arts. In addition, evidence on the effectiveness of embedding CT skills in developing domain-general CT has been inconsistent. Some studies found that explicit CT instruction within subject matter domains is an effective way of developing domain-general CT skills (e.g., Bensley and Spero 2014; Dwyer et al. 2012; Solon 2007), whereas several others reported an insignificant effect (e.g., Anderson et al. 2001; McLean and Miller 2010; Toy and Ok 2012). Furthermore, it is unclear whether instructional intervention that aims to promote domain-general CT skills also improves students' ability to solve domain-specific CT tasks, and vice versa (Fischer et al. 2014; Siegel 1988). In view of the dearth and inconsistency of the existing empirical evidence, determining the features of instructional interventions that contribute to developing domain-specific and domain-general CT remains an important challenge in CT research.

Recent developments in cognitive psychology have influenced instructional design in various ways (Elen 1995; Jonassen 1991; Merrill 2002; van Merriënboer 1997). One of the influences has been on the conception of learning and instruction. Cognitive psychologists view learning as an active, cumulative, constructive, goal-oriented, self-regulated, and situated process of knowledge and meaning building (e.g., Elen 1995; Shuell 1986; van Merriënboer 1997). Instruction is viewed as a set of activities that aim to support and enable learning, and that means helping and guiding students to actively process information, monitoring their performance, and providing feedback with respect to the appropriateness of students' learning activities (Elen 1995; Merrill 2013). These moderate constructivist views on learning and instruction (Elen 1995) emphasize that learning and understanding go hand in hand (e.g., Shuell 1986). Echoing this view, Perkins and Unger (1999) argued that understanding a subject matter domain is a matter of being able to think critically and act competently with one's knowledge of the subject matter. This implies that meaningful subject matter learning in any domain inherently involves the development of relevant CT skills. From this follows the development of CT is essentially an implicit goal in all subject matter learning.

Despite the theoretical claim that subject matter instruction in any domain can stimulate the development of CT (Perkins and Salomon 1989; Resnick et al. 2010; Resnick 1987; Smith 2002; van Merriënboer 1997), the potential impact of the design of subject matter instruction has been overlooked in existing CT research. The development of CT is largely explored through loosely defined instructional interventions that consist of teaching general CT skills within less optimally designed subject matter instruction (Tiruneh et al. 2014). Research attempts to embed CT skills within subject matter instruction have not systematically built on instructional design research, and the link between the acquisition of domain-specific and domain-general CT skills appears to be vague. In sum, although it is unclear to what extent systematically designed subject matter instruction in itself promotes the development of domain-specific and domain-general CT skills, strong impact on the development of domain-specific CT skills is to be expected since they are an integral part of the domain-specific expertise that instruction aspires toward.

Drawing on past research on cognitive development (Glaser 1984; Perkins and Salomon 1989), we explored the question of whether systematically designed subject matter instruction may facilitate the acquisition of domain-specific and, to some extent, domain-general CT skills. The aim of this paper is therefore to examine the effectiveness of systematically designed subject matter instruction in promoting the development of domain-specific and domain-general CT skills, and to investigate the relationship between the two.

Teaching CT in higher education: state-of-the-art

What is CT?

Existing literature suggests widespread disagreement among educators and researchers with regard to the definition of CT and what is to be accomplished in teaching it. Ennis (1993) defines CT as logical and reflective thought that focuses on a decision on what to believe or do. Halpern (1998, 2014) defines CT as the use of thinking strategies that increase the probability of a desirable outcome. Together with her definition, Halpern identified five major categories of CT skills: verbal reasoning, argument analysis, hypothesis testing, likelihood and uncertainty analysis, and decision-making and problem-solving. Halpern argues that the use of CT skills in solving various cognitive tasks can increase the probability of ‘a desirable outcome’ (Halpern 1998, p. 450). McPeck (1990) defines CT as the appropriate use of reflective skepticism within the problem area under consideration, and he closely relates the problem areas to particular subject matter domains.

Some researchers (Facione 1990a; Halpern 1998; Norris 1989; Perkins et al. 1993) have moreover argued that in addition to mastery of a set of cognitive skills, a more meaningful and comprehensive understanding of CT must include CT dispositions. The latter refers to a person’s inclination to use CT skills appropriately without prompting, and with conscious intent in a variety of settings, for instance, when faced with problems to solve, ideas to evaluate, or decisions to make (Ennis 1993; Halpern 1998). Researchers have arrived at a list of CT dispositions that in the main includes open-mindedness, inquisitiveness, systematicity, analyticity, truth-seeking, self-confidence, and maturity (Facione 1990a). Halpern (1998) also notes that a critical thinker demonstrates the following dispositions:

- (a) willingness to engage and persist in a complex task, (b) habitual use of plans and the suppression of impulsive activity, (c) flexibility or open-mindedness, (d) willingness to abandon non-productive strategies in an attempt to self-correct, and (e) awareness of the social realities that need to be overcome (such as the need to seek consensus or compromise) so that thoughts can become actions. (p. 452).

We used Halpern’s (2014) classification of CT skills for the purposes of this study. After synthesizing the various conceptions of CT (e.g., Bailin et al. 1999; Ennis 1989; Halpern 2014; McPeck 1990; Resnick et al. 2010; Smith 2002), we defined CT as the proficiency a person demonstrates in using thinking strategies to accomplish a task in a reasonable manner. The thinking task in question may require specific subject matter expertise for it to be reasonably performed, and we call such proficiency domain-specific CT. On the other hand, the thinking task in question may not require specific subject matter expertise, but rather knowledge of everyday life. We refer to such proficiency as domain-general CT.

Specificity and generality of CT and its implications for instruction

The question of whether CT is a set of general skills that can be applied across domains or whether it is by and large specific to a particular domain has been the subject of heated debate (e.g., Bailin et al. 1999; Davies 2013; Ennis 1989; McPeck 1990; Moore 2011; Norris 1989; Paul 1985). This disagreement has had major implications for approaches to integrate CT in higher education curricula. Generalists (Davies 2013; De Bono 1991; Ennis 1989; Halpern 1998; Kuhn 1999) claim a set of CT skills exists that are general and

applicable across a wide variety of domains. They contend that this set of general CT skills can be taught either as a specific curriculum subject (i.e., a stand-alone course), or be integrated explicitly into regular courses. On the other hand, specifists (McPeck 1990; Moore 2004, 2011) argue that thinking is highly dependent on specific domain knowledge and that CT teaching should therefore always be pursued within the context of a specific domain. McPeck (1990) has strongly argued against the notion of general CT skills on the basis that the thinking skills required in one domain are different from those required in another. This specifist position implies that each domain will need to identify its own distinctive thinking skills, and students will learn those domain-specific CT skills while building up knowledge of that particular domain.

However, it seems that the generality versus specificity debate has recently shifted towards a synthesis of the two views (Davies 2013; Robinson 2011; Smith 2002). First, although the related content and issues differ from one domain to the next, a set of CT skills that are applicable across a wide variety of domains exists. Second, the ability to think critically on a particular task is understood to be highly dependent on knowledge of the task at hand as well as knowledge of relevant CT skills. This implies that effective CT instructional approaches need to target students' in-depth understanding of a domain and that of the relevant CT skills.

CT assessment

In tandem with the absence of a consistent CT definition, one of the main challenges in CT research has been the lack of uniform CT tests. Researchers have employed various kinds of CT tests that use a broad range of formats, scope, and psychometric characteristics to measure CT outcomes (for reviews, see Ennis 1993; McMillan 1987; Tiruneh et al. 2014). Some of the available standardized domain-general CT tests include the Cornell Critical Thinking Test (CCTT: Ennis et al. 1985), the California Critical Thinking Skills Test (CCTST: Facione 1990b), the Watson–Glaser Critical Thinking Appraisal (WGCTA: Watson and Glaser 2002), the Ennis–Weir CT Essay Test (Ennis and Wier 1985) and the Halpern Critical Thinking Assessment (HCTA: Halpern 2010). These domain-general CT tests use content from a variety of real-life situations with which test takers are assumed to already be familiar.

Except for the Ennis–Weir CT Essay test and HCTA, all the above-mentioned tests use forced-choice format items, which have been criticized for not efficiently measuring significant CT features such as drawing warranted conclusions, analyzing arguments, making decisions and systematically solving problems (Norris 1989). The HCTA is the only standardized measure of domain-general CT proficiency that uses two different types of item formats: forced-choice and constructed-response formats. Halpern claims that the constructed-response format of the HCTA measures CT dispositions (Halpern 2013).

A couple of domain-specific CT tests also exist in the science domain. The Lawson's Classroom Test of Scientific Reasoning (CTSR) is the most commonly administered test in the domain of science focused on measuring general scientific reasoning skills (Lawson 1978, 2004). It is a multiple-choice test that measures scientific reasoning skills that include probabilistic reasoning, combinatorial reasoning, proportional reasoning and controlling of variables in the context of scientific domains (Lawson 1978). Respondents do not necessarily need to have expertise in a specific science domain, rather the test focuses on general science-related issues that students can reasonably be presumed to have acquired in specific science subjects. The test mainly targets junior and senior high school students, but it is also used to assess scientific reasoning skills among college science

freshmen (Lawson 1978, 2004). The other domain-specific CT test is the biology critical thinking exam (McMurray 1991). It is a multiple-choice test with 52 questions that aims to measure university students' CT skills in biology. The Critical Thinking in Electricity and Magnetism test (CTEM) is a domain-specific CT test that was recently developed and that aims to measure students' ability to draw valid inferences, analyze arguments, solve problems, make predictions, and analyze probabilities and assumptions with respect to thinking tasks that are specific to a freshman physics course (De Cock et al. 2015). The CTEM test consists of 20 items, two of which are forced-choice; the remaining are constructed-response format items. The items were designed to mirror the five CT structural components identified in the HCTA (Halpern 2010), and target the content of an introductory electricity and magnetism course. The CTEM test was validated to prompt students' ability to demonstrate the aforementioned domain-specific CT skills.

Despite the existence of a few domain-specific CT tests, the assessment of CT has thus far mainly focused on domain-general CT skills. CT has mainly been linked with everyday problem solving, and there is a general lack of experience among researchers and educators when it comes to testing for domain-specific CT skills. As discussed in the previous section, the embedded approach aims to teach desired CT skills as part of subject matter instruction. This approach is expected to result in the acquisition of both domain-specific and domain-general CT skills. Standardized tests that measure students' ability to think critically on issues and problems that are specific to a subject matter domain, however, were hardly ever administered in the various studies that adopted an embedded approach (for review, see Tiruneh et al. 2014).

Embedding CT within regular courses: instructional approaches

Ennis (1989) divided the various approaches to embedding CT within subject matter domains into two types: *Infusion* and *Immersion*. In the Infusion approach, students are explicitly trained on how to apply CT skills as part of a specific subject matter domain instruction. Students are explicitly introduced to the desired CT skills and extensively engaged in domain-specific classroom activities that call for the application of the desired CT skills. The Immersion approach, however, aims to help students acquire the desired CT skills as they construct knowledge and skills of a subject matter domain, without explicit instruction *about* desired CT skills. The main assumption behind this approach is that proficiency in CT is by definition targeted in meaningful subject matter learning; it follows that students can learn relevant and transferrable CT skills when immersed in well-designed subject matter instruction (e.g., McPeck 1990). Given the limited empirical evidence on the effectiveness of well-designed subject matter instruction on the development of domain-specific and domain-general CT skills, the effect of an Immersion-based instructional intervention is the focus of the present study.

The present study

The central question in CT instruction appears to be identifying theoretically sound and empirically valid instructional design principles that foster the development of the desired CT skills (Perkins and Salomon 1989; van Merriënboer 2013). There are a few instructional design models that offer specific guidelines to develop learning environments that enable students to acquire complex cognitive skills. The First Principles of Instruction model is one of the instructional design models that offer explicit guidelines to designing learning environments that can promote the active and constructive acquisition of higher-

order learning outcomes (Merrill 2002, 2013). The model is a synthesis of the various instructional design models that emerged from research on the acquisition of subject matter knowledge and skills. Merrill systematically reviewed the different instructional design principles that claim to be empirically valid and abstracted five interrelated prescriptive instructional design principles: activation, demonstration, application, integration and problem-centeredness. This model emphasizes that subject matter instruction designed on the basis of those principles can result in effective, efficient and engaging learning that leads to students' acquisition of knowledge and skills that are necessary to complete complex real-world tasks (Merrill 2013).

Because of its comprehensiveness and strong theoretical foundation, the First Principles of Instruction model was chosen to guide the design of the learning environment for this study. No previous study, to our knowledge, has tested the efficacy of this model in designing instructional interventions that target the development of CT skills. A brief explanation of the First Principles of Instruction model and its implications for designing subject matter instruction is offered in the next section. A learning environment in the context of a freshman physics course was designed based on the model. The following research questions are addressed: (a) What is the effect of systematically designed subject matter instruction on the development of domain-specific CT skills? (b) What is the effect of systematically designed subject matter instruction on the development of domain-general CT skills? and (c) What is the relationship between performance on domain-specific and domain-general CT tests? In line with existing theoretical literature (e.g., Perkins and Unger 1999; Resnick et al. 2010), we hypothesized that subject matter instruction systematically designed according to the First Principles of Instruction model would produce a significantly higher acquisition of domain-specific and domain-general CT skills than regular subject matter instruction.

Method

Participants

The study participants were first-year students with physics majors at two universities in northwest Ethiopia. Students at one of the universities formed the experimental group ($n = 45$), while those at the other university constituted the control group ($n = 44$). The experimental group was comprised of 24 women and 21 men between the ages of 19 and 23 years ($M = 20.09$, $SD = .93$), while the control group consisted of 23 women and 21 men between the ages of 19 and 24 years ($M = 20.32$, $SD = .98$).

Design and development of the Immersion-based instructional intervention

The intervention focused on a freshman introductory physics course, namely introductory electricity and magnetism (E&M). At both universities, this course was taught based on a harmonized national curriculum, with the same content and credit hours. The targeted course was taught during the second semester of the 2013/2014 academic year. The intervention focused only on the first five chapters of the course: electric field, electric flux, electric potential energy, capacitor and capacitance, and direct current circuits (as specified in the course textbooks of the two universities).

In recognition of the complex and multidimensional nature of CT, an effort was first made to acquire clearer understanding of the desired CT outcomes learners ought to demonstrate after the intervention. The CT skills that were the focus of our intervention were reasoning, argument analysis, hypothesis testing, likelihood and uncertainty analysis, and problem solving and decision-making. The targeted CT skills were split into sub-skills before the instructional intervention was designed. A more precise description of each of the domain-specific and domain-general CT outcomes was subsequently developed with respect to the post-intervention performance (see Table 1). Such an in-depth analysis of the CT outcomes that we wished our students to demonstrate helped us decide on the specific and relevant instructional strategies that should be targeted while the learning environment is designed and implemented.

After the desired CT outcomes were identified, the next important phase was designing a learning environment based on the First Principles of Instruction model. Table 2 offers a brief description of the principles, the implications for instructional design, and brief examples of what happened in the actual design and implementation phase of the learning environments. Two regular course instructors from the experimental university, two physics professors, one instructional psychology professor and one doctoral candidate collaborated in designing the experimental learning environment. Efforts were made to embrace the desired CT skills as part of the regular domain-specific classroom activities during this design process.

Implementation of the experimental and control interventions

Students in both the experimental and control conditions learned the same five chapters. The lessons were taught by regular instructors at the two universities. Two instructors (one as a main instructor and the other as an assistant for the tutorial sessions only) participated in the study at each university. In order to control for the teacher effect, we involved instructors from the two universities who had the same education level (all MSc in Physics) and similar years of teaching experience.

Training the experimental instructors

The two regular instructors received adequate training to be able to teach the experimental class. Their collaboration began during the design phase of the intervention, and they were both fully informed on the purpose of the intervention and what was required of them in implementing the designed lessons. For example, we initially asked them to comment on a draft version of the lessons designed for chapter one and both instructors provided useful feedback. Their involvement and feedback continued throughout the design process of the five chapters. On a number of occasions, they reported that some of the activities and questions in the draft versions were unclear or less relevant for the targeted students. A number of modifications were accordingly made.

Moreover, to facilitate implementation of the lessons as designed and provide the necessary theoretical knowledge base, the first author and the two experimental instructors participated in 5 h of face-to-face discussions over a period of 3 days. The instructors were briefed on the overall goal of the instructional intervention as well as the specific designed lesson activities of the full five chapters.

Table 1 Description of desired domain-specific and domain-general CT outcomes

CT skill	Domain-specific CT outcome	Domain-general CT outcome
Reasoning	In the context of E&M, the student will be able to: Evaluate the validity of data Recognize errors of measurement Interpret the results of an experiment	In the context of everyday situations, the student will be able to Recognize ambiguity of terms Identify vague ideas/terms Evaluate/analyze ideas from different perspectives Use questioning and paraphrasing to comprehend text
Thinking as hypothesis testing	Identify important relationships Draw valid inferences from a given tabular or graphical information Examine the adequacy of observations/samples/repetitions of an experiment to draw a conclusion Check for adequate sample size and possible bias in sampling when making a generalization	Identify cause and effect relationships Recognize the need for more information in order to make valid conclusions Draw valid inferences from given information (tabular, graphical, expository...) Examine the adequacy of observations/samples/repetitions before drawing a conclusion
Argument analysis	Identify key parts of an argument on issues related to E&M Judge the credibility of an information source Infer a correct statement from a given data set Criticize the validity of generalizations drawn from the results of an experiment Identify relevant information that is missing from an argument	Identify key parts of an argument: e.g., given a conclusion, identify the reason(s) that support the conclusion. Provide an opinion, a reason and conclusion on issues related to daily life. Infer a correct statement from a given data set Criticize the validity of generalizations
Likelihood and uncertainty analysis	Predict the probability of events (but understand the limits of extrapolation) Identify assumptions (e.g., recognize what assumptions have to be maintained in generalizations drawn from experimental results) Understand the need for additional information in making decisions Make valid predictions	Understand the probability and likelihood of an event occurrence Identify assumptions Make valid predictions: what-if questions Understand the need for adequate sample sizes
Problem-solving and decision-making	Identify the best among a number of alternatives in solving E&M-related problems Examine relevance of procedures in solving scientific problems Evaluate solutions to an E&M-related problem Make sound, evidence-based decisions Use analogies to solve E&M-related problems	Identify/choose the best option from a number of alternatives in solving everyday problems Decide on the validity of a particular scientific explanation when applied to new situations Examine the relevance of the procedures in solving problems Use analogies to solve problems Develop reasonable, creative solutions to a problem

Experimental condition

The developed lessons were taught during the regular lecture hours. Students were divided into 10 groups of 4 or 5 students. Efforts were made to have groups that were evenly spread in terms of gender and academic performance (with the latter based on students' GPA in

Table 2 Comparison between the experimental and control learning environments in relation to the First Principles of Instruction model

Instructional principle	Description of the principle	Design and implementation of the experimental learning environment	Design and implementation of the control learning environment
Problem-centered principle	Learning is promoted when learners acquire knowledge and skills in the context of real-world problems. Problems need to be comprehensive, challenging and representative of the problems learners will encounter in real life	For each chapter, relatively complex, meaningful and comprehensive problems were carefully designed by seeing each chapter as a mini-course (based on the suggestion by Merrill 2013). An attempt was made to keep the tasks relevant to the lives of students (see Fig. 2 for a sample whole-task) and thus make them more motivating. A whole-task for a particular chapter was given one or two days before instruction began; students were subsequently asked to answer the questions in the whole-task by referring to the course textbook or consulting experts (or senior students with physics majors)	Instruction was primarily 'topic-centered'. At the beginning of a new chapter, the instructor presented information related to that chapter (or subtopic). Students were sometimes shown solutions to one or two textbook problems related to the newly presented information. At the end of the lesson, students were given selected textbook problems as homework assignments. Overall, the lessons were not designed to echo real-world problems. Comprehensive problems with real-world significance that might prompt students' CT skills were not introduced at the beginning of a chapter
Activation principle	Learning is promoted when learners activate existing knowledge and skills as a foundation for new knowledge and skills. Instruction should not begin from abstract representations that learners require more background in	Various activities that helped learners make meaningful connections between newly acquired and their prior knowledge were carefully prepared in advance and implemented during instruction. For example, learners received questions about a specific topic that aimed to relate the concepts of the new topic to their prior knowledge, and they were required to share their answers with other learners (peer sharing)	There were no systematic or adequate attempts to activate learners' prior knowledge before information on a new topic was presented. When a lesson on a new topic began, the instructor usually started by briefly explaining the topic and subsequently presenting detailed information on the topic. The instructor sometimes began a new lesson by revisiting the content of a previous lesson. In most cases, the instructor himself revisited the previous lesson rather than asking students to do this. Sometimes, the instructor encouraged students to tell him what they remembered of the previous lesson, but no further prompts were offered to help students describe the preceding lessons in detail

Table 2 continued

Instructional principle	Description of the principle	Design and implementation of the experimental learning environment	Design and implementation of the control learning environment
Demonstration principle	Learning is promoted when learners observe a demonstration of the knowledge and skills to be learned that is consistent with the type of content being presented. This principle pertains to presenting necessary concepts and facts of the subject matter domain, and modeling how the newly presented concepts and facts could be applied in solving various tasks	We mainly changed the textbook's standard numerical E&M problems into more qualitative/conceptual problems. An attempt was made to qualify the tasks so that the desired CT skills could implicitly be modeled by the instructor when presenting the solutions (e.g., what can we conclude from the answer? what other options are there to solve this problem?, what other information do we need to solve this problem?, etc.)	A great deal of information was presented, but was on telling rather than both telling and showing the information. After he introduced a lesson, the instructor presented detailed information on the topic, but he did not adequately show how the presented information might be used to solve a new problem. Tasks that might have facilitated demonstration of the newly presented information were not systematically designed in advance
Application principle	Learning is promoted when learners apply newly acquired knowledge and skills on tasks that are consistent with the type of content being taught. Learning from an application is effective when learners receive corrective feedback and are properly coached, with the coaching gradually withdrawn with every subsequent problem	Relevant and challenging E&M tasks were designed that created multiple opportunities for the students to engage in applying newly presented information. When students were engaged in solving problems, activities that facilitated instructor coaching and guidance were clearly described and implemented. For example, the instructors provided partial solutions, halted at each group and observed students' discussions, provided hints as needed, acted as group members and asked thought-provoking questions, encouraged students to formulate questions using specific verbal prompts, and facilitated discussion among group members	Students mostly listened to the instructor and took notes. They were not engaged in applying the newly presented information to solve new and meaningful E&M problems; rather the instructor gave them homework assignments to practice solving the traditional end-of-chapter problems. Moreover, there was no dedicated time for students to practice solving as many practical and comprehensive questions as possible during the lessons. Even when they were asked questions, the questions focused on recalling information and did not invite further elaboration and explanations from the students. Group activities took place during some of the sessions. However, the activities for small group activities were not adequately and purposely designed. The instructor did not adequately coach the group activities and feedback was limited

Table 2 continued

Instructional principle	Description of the principle	Design and implementation of the experimental learning environment	Design and implementation of the control learning environment
Integration principle	<p>Learning is promoted when learners integrate their new knowledge into their everyday lives by being required to reflect on, discuss or defend their newly acquired knowledge or skills via peer collaboration and peer critique.</p> <p>Instruction should provide learners an opportunity to reflect on how the newly acquired knowledge and skills relate to what they already know, to explain the learned knowledge and skills to others, and to defend what they know when challenged</p>	<p>Activities that encourage students to present their solutions either to group members or full class were designed, and both peer and instructor feedback was offered. At the end of each chapter, a two-hour tutorial session was organized. The sessions mainly focused on revising the main topics of each chapter by asking students to prepare a summary (e.g., by using concept maps) of the facts and concepts discussed in the chapter, and solving a few E&M problems. Students were required to attempt to solve all the problems in advance. During these sessions, students were asked to discuss their solutions in their respective groups, and the tutors acted as coaches during the group activities. Representatives from at least two groups were asked to present solutions to a particular question in front of the full class. Students in other groups were encouraged to ask questions, and the student presenters were asked to defend their solutions when challenged by their classmates or the instructors</p>	<p>Students usually did not have the opportunity to present and defend their solutions to the full class. Interaction between the students during the lessons was very limited: they did not engage in exchanging ideas and explaining solutions to problems between themselves or to the instructor. At the end of each chapter, a two-hour tutorial session was arranged so that students could solve exercises in groups. The regular instructor and his assistant provided assistance to the students during the tutorial sessions. In most cases, however, the tutorial questions did not encourage students to apply what they had learned to solve new and meaningful problems. The questions usually promoted retention of information</p>

the first semester). Students received guidance in performing both the individual and group activities that had been designed. At the beginning of each chapter, students were assigned contextually relevant E&M problems that required them to collaborate to find solutions. Throughout the intervention, students were made to observe well-scripted instructor demonstrations that modeled the important procedures and reasoning involved in solving various E&M problems. The demonstrations were followed by extensive opportunities for the students to practice solving E&M problems both individually and in small groups for a substantial amount of time. A number of activities that encouraged students to activate prior knowledge and communicate their ideas to both their group and the entire class were carefully designed and implemented. Both peer and instructor feedback was provided as needed. Overall, students were carefully assisted in developing an in-depth understanding

of the subject matter domain, and they were coached and supported in the acquisition of the CT outcomes through the various domain-specific instructional activities. The first author monitored overall implementation of the intervention, which lasted 8 weeks. Three lessons of 2 h each were taught every week. See Table 2 for a brief overview of the activities designed and implemented in the experimental class.

Control condition

Students in the control condition followed the regular subject matter instruction. Two instructors (one main instructor and one assistant for only the tutorial sessions) from the control university were responsible for designing and implementing the lessons. The lesson durations for this group were the same that for the experimental group: a total of 8 weeks with 3 lessons of 2 h each per week. This group was similar to the experimental group in terms of previous course and parallel courses enrollment during the intervention. However, the E&M lessons for this group were not designed according to the First Principles of Instruction model, and we will refer to the instructional method in the control class as “regular” E&M instruction. See Table 2 for a detailed comparison of the control and experimental learning environments. To obtain an overview of the instructional processes, the first author observed one of the control group’s lessons. In addition, interviews were conducted with the E&M instructor on three separate occasions (at the beginning of the semester, a month after the semester, and at the posttest) to acquire additional information on the various classroom activities. A brief description of the instructional activities that took place in the control group is offered below.

At the beginning of each chapter, the main instructor gave a brief overview of the general learning outcomes. He immediately proceeded by discussing the first subtopic of a chapter and asked oral questions between presentations that encouraged students to engage in discussions. However, students were not pushed to give more detailed explanations of their responses. In most cases, the instructor himself offered the explanations. He usually showed the solutions to one or two problems after a brief discussion of a particular topic. In most cases, students took notes and wrote down the solutions. Towards the end of the lesson, students were usually handed homework that was to be solved by the next lesson. The students, however, did not receive comprehensive and contextually relevant E&M tasks at the beginning of each chapter. The E&M problems solved by the teacher during class and those given as homework assignments were traditional end-of-chapter problems that focused on computation and gave students limited opportunities to engage in thoughtful discussions (see Fig. 1 for a comparison of E&M problems for the control and experimental conditions).

Instruments

The effects of an instructional intervention on the development of CT skills should be measured by using valid and reliable CT measures that are sensitive enough to capture the changes of targeted CT outcomes (Ennis 1993; Halpern 1993; McMillan 1987). The CTEM test was administered in order to measure students’ acquisition of the desired domain-specific CT outcomes. The HCTA (Halpern 2010) was administered to measure the acquisition of domain-general CT outcomes. A pilot study was conducted to examine the applicability of the HCTA for use to the present participants. The test consists of 25 scenarios (5 scenarios for each domain-general CT skills targeted in the study), with variety of everyday health, education, politics and social policy issues. Each scenario is

<i>Sample E&M problem in the control class</i>	<i>Sample E&M problem in the experimental class</i>
<p>A parallel plate capacitor has a square plate of side 10cm, and separation 4mm. A dielectric slab of dielectric constant $k = 2$ has the same area as the plates but has a thickness of 3mm. What is the capacitance (a) without the dielectric, and (b) with the dielectric?</p>	<p>You may be aware that many of the standard computer keyboard buttons are constructed of capacitors. The keys are spring-loaded and each key has a tiny plate attached to the bottom of it. When you press a key, it moves this plate closer to the plate below it. So, basically, when a key is pushed down, the soft insulator between the movable plate and the fixed plate is compressed, changing the capacitance. This change in capacitance helps the computer to recognize which key is pressed.</p> <p>Let us assume that the separation between the plates is initially 5.00 mm, but is reduced to 0.150 mm when a key is pressed. The plate area is $9.50 \times 10^{-5} \text{ m}^2$ and the capacitor is filled with a material whose dielectric constant is 3.50. Determine the change in capacitance detected by the computer. Explain the relationship among plate area, dielectric material, and capacitance.</p>

Fig. 1 Sample E&M problems for the control and experimental condition

followed by questions that require respondents to provide a constructed response and to subsequently select the best option from a short list of alternatives (forced-choice items). Based on the findings of the pilot study, 5 scenarios (1 from each CT category) that were somewhat confusing and reduced the test's overall internal consistency in this particular context were omitted. As a result, 20 constructed-response and 20 forced-choice items were ultimately administered.

Both the CTEM and HCTA focus on similar CT components, with the exception that the CTEM items focus on E&M tasks, while the HCTA items focus on thinking tasks drawn from everyday life that do not require specific subject matter expertise (see Fig. 3

<p>You have probably heard about the risk of lightning strike on human life. During a stormy day, your parents may have advised you not to walk outside on the street so that you may not encounter an electric shock from lightning strikes. On the other hand, you know that flying in an airplane during a stormy weather is completely safe with respect to the electric shock that you might encounter due to lightning. You know that most aircrafts skin consist primarily aluminum, which conducts electricity very well. In addition, you have been hearing that airplanes experience lightning strikes during flight, but apparently the electric shock from the lightning is not felt by passengers inside the plane.</p> <p>Although the external part of the aircraft conducts electricity, you have come to know that the electric current from the lightning strike remain on the exterior of the aircraft. Why do you think the presence of this electric charge on the external body of the airplane is not felt within the airplane? What explanations do you give to this phenomenon?</p>

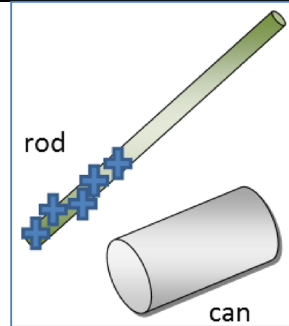
Fig. 2 Sample whole-task for chapter three

Sample CTEM item:

Hanna does the following experiment: she brings a positively charged rod close to a metal can. Doing the experiment shows that the can is attracted to the rod.

Hanna is puzzled with the result of her experiment. She expected the negative electrons would be attracted to the rod while the positive nuclei are repelled, and opposite forces cancel out, which would mean that the can remains at rest.

How can you make Hanna's argument consistent with the experiment? Give an explanation.

*Sample HCTA item:*

After a televised debate on capital punishment, viewers were encouraged to log on to the station's web site and vote online to indicate if they were "for" or "opposed to" capital punishment. Within the first hour, almost 1000 people "voted" at the website, with close to half voting for each position. The news anchor for this station announced the results the next day. He concluded that the people in this state were evenly divided on the issue of capital punishment.

Given these data, do you agree with the announcer's conclusion? ☐ Yes ☐ No

Provide two suggestions for improving this study:

First suggestion: _____

Second suggestion: _____

Fig. 3 Sample CTEM and HCTA items

for sample CTEM and HCTA items). We computed the internal consistencies (Cronbach's alpha) of the administered tests in the present study: .74 for the CTEM, .76 for the HCTA constructed-response, .73 for the HCTA forced-choice and .77 for the HCTA overall test. Although a desirable value for internal consistency may vary as a function of the nature of the construct being measured, Cronbach's alpha values between .70 and .80 are considered acceptable (Cohen et al. 2007). Prior physics knowledge of the participants (physics scores from the Ethiopian Higher Education Entrance Examination) was collected from the student records offices of the two universities.

Procedure

The CTEM was administered as a posttest-only test a week after the end of the intervention. Because the CTEM items require prior knowledge of E&M, we felt it was reasonable to administer the test only at the end of the intervention. The HCTA test, on the other hand, was administered both to the experimental and control groups as a pretest, immediately before the beginning of the intervention and as a posttest a week after the end of the intervention. Due to practical reasons, the paper version of the HCTA test was administered since computer-based administration of the HCTA

was not possible. Participants were required to first answer all the constructed-response format items and then the forced-choice format items. Administration of the CTEM test lasted between 60 and 75 min, and the HCTA (both formats) between 70 and 90 min.

Approximately 90 % of the experimental lessons were observed, and the experimental instructor was consulted after each lesson to reflect on challenges that surfaced as well as any other aspects that might improve implementation of the lessons as designed. Post-lesson discussions focused on such issues as usage of instructional time, giving of support and feedback to groups within the allocated instructional time, oral questions used to prompt students to further elaborate on their answers, and overall evaluation of the implementation of the lesson in relation to the design. Instructors registered class attendance for each session both in the experimental and control conditions. Eighty-five percent of the experimental group students and approximately 80 % of the control group students attended more than 90 % of the sessions. There were two dropouts in the experimental group and one dropout in the control group. The pretest data of those three students were omitted from the results. This means that our analysis of the data from the two groups is based on 45 students for the experimental group and 44 students for the control group.

Results

Screening of the data

The CTEM and HCTA scores were screened for accuracy of data entry, missing values and the assumptions for normality and homogeneity of variances. A separate overview of the experimental and control students' scores for each CTEM and HCTA items showed random missing data for a few items. However, the proportion of missing values per item was very limited (<5 %) and scattered over each of the 20 CTEM and HCTA items. Mean substitution was therefore used to estimate the missing data. The mean scores for each separate item for the experimental and control groups were calculated and the handful missing values were substituted with the respective group mean scores. Outliers were also separately sought in the experimental and control groups. Visual inspection of boxplots and inspection of the z scores for each of the CTEM and HCTA variables showed that there were no potential outliers.

Moreover, tests of assumptions for normality and homogeneity of variances were conducted through examination of the standardized residuals for the CTEM and HCTA scores. For the CTEM, a Shapiro–Wilk's test ($p > .05$) and a visual inspection of the histograms, the Q–Q plot and boxplot suggested that the scores from the two groups were approximately normally distributed. Using the standardized residuals, the assumption of homogeneity of variances was tested and satisfied based on Levene's F test, $F(1, 87) = 1.57$, $p = .11$. For the HCTA scores, a Shapiro–Wilk's test ($p > .05$) and a visual inspection of the histograms and boxplot showed that the HCTA pretest and posttest scores were also approximately normally distributed for both the experimental and control groups. Furthermore, the assumptions of homogeneity of variances were tested and satisfied based on Levene's F test for the pretest ($F(1, 87) = .16$, $p = .69$) and posttest scores ($F(1, 87) = 1.36$, $p = .25$).

Domain-specific CT performance: CTEM

Initial comparison of prior physics knowledge revealed no significant differences between the experimental and control group, $t(87) = .15$, $p = .88$. An independent sample t test was therefore conducted to compare the performance of the two groups on the domain-specific CT test. The results indicated that the CTEM mean score for the experimental group was significantly higher than that of the control group, $t(87) = 7.15$, $p < .001$, $d = 1.55$. The effect size for this analysis was found to exceed Cohen's (1988) convention for a large effect ($d = .80$).

An analysis of covariance (ANCOVA) was conducted to examine whether the statistically significant mean score differences could be maintained after controlling for physics prior knowledge. The ANCOVA results showed that the CTEM mean score of the experimental group was significantly higher than that of the control group, $F(1, 86) = 52.56$, $p < .001$, $\eta^2 = .379$. The results indicated that the intervention accounted for 37.9 % of the variance in the acquisition of domain-specific CT skills. Post-hoc power analysis by using G*Power (Faul et al. 2007) indicated that the power to detect the effect size observed in the present study ($d = 1.55$, $p < .001$) was $>.99$. The a priori power analysis indicated that a total sample size of 84 would be sufficient to detect a large effect ($d = .8$; Cohen 1988) with a power of .95 ($p = .05$), and a total sample size of 210 would be sufficient to detect a medium effect ($d = .5$; Cohen 1988) with a power of .95 ($p = .05$). See Table 3 for descriptive statistics of the CTEM test.

Domain-general CT performance: HCTA

In order to examine the effect of the instructional intervention on students' domain-general CT performance, a 2 (groups: experimental and control) \times 2 (testing time: pretest and posttest) mixed design ANOVA was conducted. The results of the mixed design ANOVA revealed that the two groups together demonstrated a statistically significant improvement on the HCTA mean scores across the two time points, $F(1, 87) = 4.61$, $p = .035$, $\eta^2 = .05$. The effect size value ($\eta^2 = .05$) suggested a small practical significance. However, there was no significant interaction between the intervention type (experimental-control) and the testing time (pretest-posttest), $F(1, 87) = .14$, $p = .71$. In other words, the HCTA mean score for the experimental group did not show a significant pretest-posttest improvement compared to the control group. This indicates that the experimental learning environment did not result in a significantly greater pretest-posttest improvement in the acquisition of domain-general CT skills compared to the control learning environment. The descriptive statistics of the HCTA scores are shown in Table 3.

Table 3 Descriptive statistics for experimental and control groups: prior knowledge, CTEM and HCTA scores

Group	Prior knowledge	CTEM		Pretest HCTA		Posttest HCTA	
	M (SD)	M (SD)	Range	M (SD)	Range	M (SD)	Range
Experimental	40.93 (8.21)	31.56 (5.44)	21	71.49 (4.76)	19	73.04 (5.21)	24
Control	41.18 (7.38)	24.57 (3.57)	14	70.36 (4.11)	17	71.45 (5.12)	18

Relationship between domain-specific and domain-general CT performances

Calculation of the Pearson's correlation coefficient showed a significant positive relationship between pretest HCTA and posttest HCTA scores ($r = .29, p = .006$). Moreover, the CTEM scores significantly correlated with the posttest HCTA scores ($r = .38, p = .01$). These findings show that when both groups are taken together, those students who scored higher on the pretest HCTA also tended to score higher on the posttest HCTA. Post-intervention comparison similarly indicated that those who scored higher on the CTEM test also tended to score higher on the posttest HCTA. A linear regression analysis also revealed that the CTEM test explained a significant proportion of the variance on posttest HCTA performance, $F(1, 87) = 14.7, p = .05, R^2 = .145$. The result shows that CTEM performance was a significant predictor, accounting for 14.5 % of the variance in posttest HCTA scores. Post-hoc power analysis using G*Power (Faul et al. 2007) indicated that the power to detect the observed effect at the .05 level was .94 for the regression in prediction of the posttest HCTA performance.

Discussion

In this study, we argued that the design of CT instructional interventions should be supported by the principles of instructional design research. To that end, we tested an alternative method to address the challenge of CT development through the systematic design of subject matter instruction rather than explicit instruction on general CT skills. A regular physics course was systematically designed in accordance with the First Principles of Instruction model. We hypothesized that E&M instruction systematically designed in line with the First Principles of Instruction model would produce higher acquisition of domain-specific and domain-general CT skills than regular E&M instruction.

Implementation of the lessons for the experimental condition was carefully monitored, and sufficient information was gathered with respect to the implementation of the lessons in the control condition. With regard to the first research question, we found that a systematically designed E&M instruction that implicitly targeted CT skills in various domain-specific classroom activities resulted in higher acquisition of domain-specific CT skills compared to regular E&M instruction. We focused on the systematic design of subject matter instruction (supported by valid principles of instructional design research) as previous CT intervention studies did not systematically explore how subject matter instruction in itself may stimulate learning of domain-specific CT skills. The instructional interventions designed and implemented as part of a couple of previous Immersion-oriented CT empirical studies (e.g., Barnett and Francis 2012; Garside 1996; Renaud and Murray 2008; Stark 2012; Wheeler and Collins 2003) appear to show significant limitations. The interventions focused mainly on a specific component of the learning environment (e.g., small group discussion only), and only minimally emphasized other important learning environment components such as the types of learning tasks/problems designed for discussion (e.g., are the learning tasks challenging enough to provoke discussion among students? Are the tasks authentic/contextually relevant?). They also paid scant attention to the adequacy of support, feedback and coaching offered during full-class and small group discussions. In most previous CT studies, the desired CT outcomes learners were expected to demonstrate after instruction were moreover barely described or articulated during the design phase. It is next to impossible to evaluate the extent to which the various designed tasks and

instructional activities were relevant in stimulating the acquisition of the desired CT outcomes.

For the present study, efforts were made to design a learning environment that addressed the limitations of previous studies. First, the desired domain-specific and domain-general CT outcomes were operationalized and described. A learning environment that could stimulate the acquisition of the desired CT outcomes was subsequently systematically designed. In accordance with the theoretical claim that meaningful subject matter learning inherently involves development of relevant CT skills (e.g., Glaser 1984; Resnick 1987), the E&M instruction was systematically designed in such a way that it provided students with the opportunity to engage in a number of domain-specific classroom activities. It is important to point out that previous studies already implemented one or two of the instructional strategies implemented in the present study. For example, the discussion method of teaching (e.g., Wheeler and Collins 2003), and teacher modeling (e.g., Anderson et al. 2001) are among the most commonly employed instructional strategies in previous Immersion-oriented CT studies. However, for this study, we designed a comprehensive intervention that integrates most of the empirically validated instructional design principles. The findings with regard to domain-specific CT skills suggest that systematic design of subject matter instruction based on a combination of empirically valid instructional principles promotes the acquisition of domain-specific CT skills. CT development, this study argues, involves both domain-specific and domain-general dimensions. It demonstrates that acquisition of domain-specific CT skills can be improved through systematic design of subject matter instruction without explicit teaching of general CT skills. This finding is consistent with the result of a recent meta-analysis of strategies for teaching CT (Abrami et al. 2015) as well as previous theoretical claims (e.g., Glaser 1984; McPeck 1990; Resnick et al. 2010; Resnick 1987) that underlined the importance of learning environments systematically designed in accordance with relevant instructional principles.

For the second research question, however, the findings showed that the experimental learning environment did not result in a statistically significant improvement for domain-general CT skills compared to the control learning environment. Gains in domain-specific CT proficiency found in the experimental condition were not accompanied by gains in domain-general CT proficiency. The two groups together demonstrated improvement in the acquisition of domain-general CT skills, between the pretest and posttest scores. The same test was administered both prior to and after the intervention, and the observed pretest–posttest improvement might simply be a test–retest effect.

On the other hand, we found that domain-specific CT proficiency significantly predicted posttest domain-general CT proficiency. This suggests that when a domain-general CT test that presumably required similar thinking skills was administered to the participants, performance on a domain-specific CT test was a significant predictor of performance on a domain-general CT test. To a degree, this reveals a tendency to transfer the acquired domain-specific CT skills in solving domain-general CT tasks. This finding is consistent with previous psychology studies in which higher performance on a psychological CT test also predicted higher performance on a domain-general CT test (e.g., Williams et al. 2004).

A number of reasons may explain why the designed learning environment did not have a significant effect on the acquisition of domain-general CT skills. The absence of an explicit focus on the desired CT skills during the E&M instruction may have kept students from abstracting the domain-specific CT skills and applying them in solving domain-general tasks. This suggests that a great emphasis on systematic development of domain-specific knowledge alone may not be sufficient to facilitate transfer of domain-specific CT skills to

everyday problems. Perhaps a worthwhile approach to CT instruction may be to explicitly emphasize desired CT skills within specific subject matter instruction. Proponents of the embedded approach often claim that explicitly teaching CT skills within subject matter instruction is the best way to stimulate development of transferrable CT skills (Davies 2013; Halpern and Hakel 2002; Halpern 1998). For example, some generalists have argued that students must be aware that they are being taught CT skills during specific subject matter instruction and they will be expected to use those skills to solve everyday problems or issues they will come across. However, the main criticism that has been directed at generalists is that they largely see CT as everyday problem-solving that is detached from domain-specific CT proficiency (see Bailin et al. 1999; Resnick et al. 2010; Smith 2002). To date, there is no agreement on how specific subject matter instruction can be optimally designed to develop both domain-specific and domain-general CT skills. An important area for future studies would therefore be to evaluate the effectiveness of explicit teaching of CT skills within well-designed subject matter instruction to develop both domain-specific and domain-general CT skills. It could prove interesting to compare an Immersion-based learning environment with an Infusion-based learning environment in which CT skills are explicitly trained within systematically designed subject matter instruction.

Another possible explanation for the insignificant effect on the acquisition of domain-general CT skills may relate to the longstanding debate around the specificity and generality of CT skills. As noted in our above analysis of existing CT literature, generalists (e.g., De Bono 1991; Ennis 1989; Siegel 1988) view CT skills as applicable across domains, whereas specifists (e.g., McPeck 1990) argue against the existence of general CT skills on the grounds that thinking always amounts to thinking about something and that specific knowledge of a subject matter is necessary for CT. In this study, students in the experimental condition were intensively engaged in acquiring deeper understanding of E&M through an implicit emphasis on the desired CT outcomes. These students performed significantly better than the control group students on domain-specific CT tasks. However, the acquired domain-specific CT proficiency did not transfer when the same students were confronted with domain-general CT tasks (viz., the HCTA). Following the specifist view, it could be argued that the study participants perhaps lacked adequate knowledge of the content used in preparing the HCTA test. This reinforces the notion that the ability to think critically is mainly content dependent (e.g., Bailin et al. 1999; McPeck 1990; Smith 2002). The findings revealed that, compared to the control group, the experimental group students were able to demonstrate proficiency in using CT skills for E&M-specific thinking tasks. However, those CT skills were not applicable when they were presented with domain-general CT tasks. Students' failure to transfer the acquired domain-specific CT skills may therefore spring from the HCTA itself. An important area for future study would therefore be to evaluate the effectiveness of CT-embedded instructional approaches through administration of at least two domain-general CT tests that were designed based on different everyday content yet focused on similar CT skills.

A third possible explanation for the unimproved domain-general CT skills may relate to the brief duration of the intervention: 8 weeks and with a focus on just 50 % of the E&M course content. Perhaps the intervention was too short to produce a substantial change in participants' modes of thought, which made it impossible for them to transfer the acquired domain-specific CT skills to other domains than the E&M problems. Moreover, the experimental group students were also simultaneously following other courses in which subject matter instruction appeared to be less systematically designed. This may have resulted in limited opportunities for students to extensively practice the desired CT skills in other subject matter domains, and hence hindered their transfer. An important implication

of this finding is that transfer of domain-specific CT skills to everyday problems may not automatically occur during a brief instructional intervention, but may instead require a conscious and systematic design of all subject matter instruction toward CT.

Study limitations

The findings of this study are based on a comparison of two intact classrooms at different universities taught by different instructors. Although the initial plan was to use two intact groups at the same university, the number of first-year students with major physics at the targeted university was very limited with just one intact group. To minimize the effects of having two different instructors and institutions, efforts were made to recruit instructors from the two universities with similar education levels and equivalent years of teaching experience. Efforts were also made to closely monitor the implementation of the lessons at both the experimental and control universities. However, it is important to interpret the findings from the present study by taking into consideration the limitations that sprang from having different institutions and instructors. Moreover, random assignment of the two intact groups into an experimental and control condition was not feasible. The first author is affiliated with one of the two universities. Since we expected to intensively collaborate with the regular instructors and to make the close follow-up more convenient, the group at the affiliated university was purposely assigned to the experimental condition.

Conclusion

This study explored the effectiveness of systematically designed subject matter instruction on the development of domain-specific and domain-general CT skills. It demonstrated that a typical freshman course systematically designed based on the First Principles of Instruction model—with an implicit focus on the desired CT outcomes as an integral part of the domain-specific classroom activities—can stimulate the development of domain-specific CT skills. This finding suggests that systematic design of subject matter instruction needs to be made an important component of teaching and learning in undergraduate education if students are to demonstrate domain-specific CT proficiency. Although this study's instructional intervention failed to provide evidence of the transfer of the acquired domain-specific CT skills to everyday problems, this does not mean that domain-general CT skills cannot be systematically taught. Our hope is that the present study will encourage researchers and instructional designers to pay attention to systematic design of subject matter instruction as a valuable approach to addressing the challenges of CT development. The following observations with regard to CT research in undergraduate education were particularly important. First, we showed that both the domain-specific and domain-general CT outcomes that we wish students to demonstrate need to be identified and precisely articulated before any attempts at teaching CT. Second, through a systematic design of regular subject matter instruction, useful empirical evidence was presented that supports the longstanding theoretical claim that meaningful subject matter learning in a domain can result in the development of domain-specific CT skills. Third, following the argument that embedding CT within subject matter domains should result in the acquisition of both domain-specific and domain-general CT skills, CTEM and HCTA tests were administered respectively to evaluate the effectiveness of the designed instructional intervention. Accordingly, empirical evidence that establishes the relationship between acquisition of

domain-specific and domain-general CT skills, a barely examined research question, was validated. Our starting point was that instructional interventions for CT are not sufficiently supported by the principles of instructional design research. Through this study, we hope to have demonstrated how the two largely detached fields of CT and instructional design research can systematically be integrated. We moreover argued that the instructional principles behind various instructional design models are not sufficiently attuned to specific instructional settings. In this study, we hope to have shown how those empirically valid instructional design principles can be translated into usable instructional design prescriptions that are also relevant to CT.

Acknowledgments We would like to thank Dawit A. Getahun and Stephanie Lem for their useful advice with respect to data analysis. We would also like to express our deepest appreciation to the anonymous reviewers for the constructive comments and suggestions on earlier versions of the manuscript.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research*, 85(2), 275–314. doi:[10.3102/0034654314551063](https://doi.org/10.3102/0034654314551063).
- Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., & Zhang, D. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research*, 78(4), 1102–1134. doi:[10.3102/0034654308326084](https://doi.org/10.3102/0034654308326084).
- Anderson, T., Howe, C., Soden, R., Halliday, J., & Low, J. (2001). Peer interaction and the learning of critical thinking skills in further education students. *Instructional Science*, 1–32. <http://link.springer.com/article/10.1023/A:1026471702353>.
- Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago: The University of Chicago Press.
- Association of American Colleges and Universities. (2005). *Liberal education outcomes: A preliminary report on student achievement in college*. Washington, DC: AAC&U.
- Bailin, S., Case, R., Coombs, J. R., & Daniels, L. B. (1999). Common misconceptions of critical thinking. *Journal of Curriculum Studies*, 31(3), 269–283. doi:[10.1080/002202799183124](https://doi.org/10.1080/002202799183124).
- Barnett, J. E., & Francis, A. L. (2012). Using higher order thinking questions to foster critical thinking: A classroom study. *Educational Psychology*, 32(2), 201–211.
- Bensley, D., & Spero, R. (2014). Improving critical thinking skills and metacognitive monitoring through direct infusion. *Thinking Skills and Creativity*, 12, 55–68. doi:[10.1016/j.tsc.2014.02.001](https://doi.org/10.1016/j.tsc.2014.02.001).
- Billing, D. (2007). Teaching for transfer of core/key skills in higher education: Cognitive skills. *Higher Education*, 53(4), 483–516. doi:[10.1007/s](https://doi.org/10.1007/s).
- Butler, H. A., Dwyer, C. P., Hogan, M. J., Franco, A., Rivas, S. F., Saiz, C., & Almeida, L. S. (2012). The Halpern critical thinking assessment and real-world outcomes: Cross-national applications. *Thinking Skills and Creativity*, 7(2), 112–121. doi:[10.1016/j.tsc.2012.04.001](https://doi.org/10.1016/j.tsc.2012.04.001).
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). London: Routledge.
- Davies, M. (2013). Critical thinking and the disciplines reconsidered. *Higher Education Research & Development*, 32(4), 529–544. doi:[10.1080/07294360.2012.697878](https://doi.org/10.1080/07294360.2012.697878).
- De Bono, E. (1991). The direct teaching of thinking in education and the CoRT method. In S. Maclure & P. Davies (Eds.), *Learning to think: Thinking to learn* (pp. 3–14). Oxford: Pergamon Press.

- De Cock, M., Tiruneh, D., Sermeus, J., & Elen, J. (2015). Development and validation of domain-specific critical thinking tests in physics. In *The Conference of International Research Group on Physics Teaching (GIREP EPEC 2015)*. Wroclaw.
- Dwyer, C. P., Hogan, M. J., & Stewart, I. (2012). An evaluation of argument mapping as a method of enhancing critical thinking performance in e-learning environments. *Metacognition and Learning*, 7(3), 219–244. doi:[10.1007/s11409-012-9092-1](https://doi.org/10.1007/s11409-012-9092-1).
- Elen, J. (1995). *Blocks on the road to instructional design prescriptions*. Leuven: Leuven Universiteit Press.
- Ennis, R. H. (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher*, 18(3), 4–10. doi:[10.3102/0013189X018003004](https://doi.org/10.3102/0013189X018003004).
- Ennis, R. H. (1993). Critical thinking assessment. *Theory into Practice*, 32, 179–186.
- Ennis, R. H., Millman, J., & Tomko, T. N. (1985). *Cornell critical thinking test level Z*. Pacific Grove, CA: Midwest publications.
- Ennis, R. H., & Wier, E. (1985). *The Ennis–Wier critical thinking essay test*. Pacific Grove, CA: Midwest publications.
- Facione, P. A. (1990a). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. Research findings and recommendations*. Newark, DE: American Philosophical Association. ERIC Document Reproduction Service No. ED315 423.
- Facione, P. A. (1990b). *The California critical thinking skills test: College level. Technical report #2. Factors predictive of CT skills*. Millbrae, CA: California Academic Press.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavioral Research Methods*, 39(2), 175–191. doi:[10.3758/BF03193146](https://doi.org/10.3758/BF03193146).
- Fischer, F., Kollar, I., Ufer, S., Sodian, B., Hussmann, H., Pekrun, R., et al. (2014). Scientific reasoning and argumentation: Advancing an interdisciplinary research agenda in education. *Frontline Learning Research*, 4, 28–45. doi:[10.14786/flr.v2i2.96](https://doi.org/10.14786/flr.v2i2.96).
- Garside, C. (1996). Look who's talking: a comparison of lecture and group discussion teaching strategies in developing critical thinking skills. *Communication Education*, 45, 212–227.
- Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist*, 39(2), 93–104. doi:[10.1037//0003-066X.39.2.93](https://doi.org/10.1037//0003-066X.39.2.93).
- Halpern, D. F. (1993). Assessing the effectiveness of critical thinking instruction. *The Journal of General Education*, 42(4), 238–254.
- Halpern, D. F. (1998). Teaching critical thinking for transfer across domains. *American Psychologist*, 53(4), 449–455. doi:[10.1037//0003-066X.53.4.449](https://doi.org/10.1037//0003-066X.53.4.449).
- Halpern, D. F. (2010). *Halpern critical thinking assessment: Manual*. Modling: Schuhfried GmbH.
- Halpern, D. F. (2013). The Halpern critical thinking assessment: A response to the reviewers. *Inquiry: Critical Thinking Across Disciplines*, 28(3), 28–39. doi:[10.5840/inquiryct201328317](https://doi.org/10.5840/inquiryct201328317).
- Halpern, D. F. (2014). *Thought and Knowledge: An Introduction to Critical Thinking* (5th ed.). New York: Psychology Press.
- Halpern, D. F., & Hakel, M. D. (2002). Learning that lasts a lifetime: Teaching for long-term retention and transfer. *New Directions for Teaching and Learning*, 89, 3–7. doi:[10.1002/tl.42](https://doi.org/10.1002/tl.42).
- Jonassen, D. H. (1991). Objectivism versus constructivism: Do we need a new philosophical paradigm? *Educational Technology Research and Development*, 39(3), 5–14. doi:[10.1007/BF02296434](https://doi.org/10.1007/BF02296434).
- Kuhn, D. (1999). A developmental model of critical thinking. *Educational Researcher*, 28(2), 16. doi:[10.2307/1177186](https://doi.org/10.2307/1177186).
- Lawson, A. (1978). The development and validation of a classroom test of formal reasoning. *Journal of Research in Science Teaching*, 15(1), 11–24. doi:[10.1002/tea.3660150103](https://doi.org/10.1002/tea.3660150103).
- Lawson, A. (2004). The nature and development of scientific reasoning: A synthetic view. *International Journal of Science and Mathematics Education*, 2(3), 307–338. doi:[10.1007/s10763-004-3224-2](https://doi.org/10.1007/s10763-004-3224-2).
- McLean, C. P., & Miller, N. A. (2010). Changes in critical thinking skills following a course on science and pseudoscience: A quasi-experimental study. *Teaching of Psychology*, 37(2), 85–90. doi:[10.1080/00986281003626714](https://doi.org/10.1080/00986281003626714).
- McMillan, J. H. (1987). Enhancing college students' critical thinking: A review of studies. *Research in Higher Education*, 26(1), 3–29. doi:[10.1007/BF00991931](https://doi.org/10.1007/BF00991931).
- McMurray, M. A. (1991). Reliability and construct validity of a measure of critical thinking skills in biology. *Journal of Research in Science Teaching*, 28(2), 183–192.
- McPeck, J. (1990). *Teaching critical thinking: Dialogue and dialectic*. New York: Routledge.
- Merrill, D. M. (2002). First principles of instruction. *Educational Technology Research and Development*, 50(3), 43–59.
- Merrill, D. M. (2013). *First principles of instruction: Identifying and designing effective, efficient, and engaging instruction*. San Francisco: Wiley.

- Moore, T. (2004). The critical thinking debate: How general are general thinking skills? *Higher Education Research & Development*, 23(1), 3–18. doi:[10.1080/0729436032000168469](https://doi.org/10.1080/0729436032000168469).
- Moore, T. (2011). Critical thinking and disciplinary thinking: A continuing debate. *Higher Education Research & Development*, 30(3), 261–274. doi:[10.1080/07294360.2010.501328](https://doi.org/10.1080/07294360.2010.501328).
- Niu, L., Behar-Horenstein, L. S., & Garvan, C. W. (2013). Do instructional interventions influence college students' critical thinking skills? A meta-analysis. *Educational Research Review*, 9, 114–128. doi:[10.1016/j.edurev.2012.12.002](https://doi.org/10.1016/j.edurev.2012.12.002).
- Norris, S. P. (1989). Can we test validly for critical thinking? *Educational Researcher*, 18(9), 21–26. doi:[10.2307/1176715](https://doi.org/10.2307/1176715).
- Pascarella, E. T., & Terenzini, P. T. (2005). *How college affects students: A third decade of research* (Vol. 2). San Francisco: Jossey-Bass.
- Paul, R. W. (1985). McPeck's mistakes. *Informal Logic*, 7(1), 35–43.
- Penningroth, S. L., Despain, L. H., & Gray, M. J. (2007). Topical articles: A course designed to improve psychological critical thinking. *Teaching of Psychology*, 34(3), 153–157. <http://www.eric.ed.gov/ERICWebPortal/detail?accno=EJ773896>.
- Perkins, D., Jay, E., & Tishman, S. (1993). Beyond abilities: A dispositional theory of thinking. *Merrill-Palmer Quarterly*, 39(1), 1–21.
- Perkins, D., & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher*, 18(1), 16–25. doi:[10.3102/0013189X018001016](https://doi.org/10.3102/0013189X018001016).
- Perkins, D., & Unger, C. (1999). Teaching and learning for understanding. In Instructional Design (Ed.), *Theories and models* (pp. 91–114). Mahwah, NJ: Lawrence Erlbaum Associates.
- Renaud, R. D., & Murray, H. G. (2008). A comparison of a subject-specific and a general measure of critical thinking. *Thinking Skills and Creativity*, 3(2), 85–93. doi:[10.1016/j.tsc.2008.03.005](https://doi.org/10.1016/j.tsc.2008.03.005).
- Resnick, L. (1987). *Education and training*. Washington, DC: National Academy Press.
- Resnick, L., Michaels, S., & O'Connor, M. (2010). How (well-structured) talk builds the mind. In D. D. Preiss & R. J. Sternberg (Eds.), *Innovations in Educational Psychology: Perspectives on Learning, Teaching, and Human Development* (pp. 163–194). New York: Springer Publishing Company.
- Robinson, S. R. (2011). Teaching logic and teaching critical thinking: revisiting McPeck. *Higher Education Research & Development*, 30(3), 275–287. doi:[10.1080/07294360.2010.500656](https://doi.org/10.1080/07294360.2010.500656).
- Shuell, T. J. (1986). Cognitive conceptions of learning. *Review of Educational Research*, 56(4), 411–436. doi:[10.3102/00346543056004411](https://doi.org/10.3102/00346543056004411).
- Siegel, H. (1988). *Educating reason: Rationality, critical thinking, and education*. New York: Routledge.
- Smith, G. (2002). Are there domain-specific thinking skills? *Journal of Philosophy of Education*, 36(2), 207–227. doi:[10.1111/1467-9752.00270](https://doi.org/10.1111/1467-9752.00270).
- Solon, T. (2007). Generic critical thinking infusion and course content learning in introductory psychology. *Journal of Instructional Psychology*, 34(2), 15. <http://eric.ed.gov/ERICWebPortal/recordDetail?accno=EJ774169>.
- Stark, E. (2012). Enhancing and assessing critical thinking in a psychological research methods course. *Teaching of Psychology*, 39(2), 107–112. doi:[10.1177/0098628312437725](https://doi.org/10.1177/0098628312437725).
- Tiruneh, D. T., Verburgh, A., & Elen, J. (2014). Effectiveness of critical thinking instruction in higher education: A systematic review of intervention studies. *Higher Education Studies*, 4(1), 1–17. doi:[10.5539/hes.v4n1p1](https://doi.org/10.5539/hes.v4n1p1).
- Toy, B. Y., & Ok, A. (2012). Incorporating critical thinking in the pedagogical content of a teacher education programme: Does it make a difference? *European Journal of Teacher Education*, 35(1), 39–56. doi:[10.1080/02619768.2011.634902](https://doi.org/10.1080/02619768.2011.634902).
- Van Gelder, T. (2005). Teaching critical thinking: Some lessons from cognitive science. *College Teaching*, 53(1), 41–46.
- Van Merriënboer, J. J. G. (1997). *Training complex cognitive skills: A four-component instructional design model for teaching technical training*. Englewood Cliffs, NJ: Educational Technology Publications.
- Van Merriënboer, J. J. G. (2013). Perspectives on problem solving and instruction. *Computers & Education*, 64, 153–160. doi:[10.1016/j.compedu.2012.11.025](https://doi.org/10.1016/j.compedu.2012.11.025).
- Watson, G., & Glaser, E. (2002). *Watson–Glaser critical thinking appraisal* (UK ed.). London: Pearson Assessment.
- Wheeler, L. A., & Collins, S. K. R. (2003). The influence of concept mapping on critical thinking in baccalaureate nursing students. *Journal of Professional Nursing*, 19(6), 339–346. doi:[10.1016/S8755-7223\(03\)00134-0](https://doi.org/10.1016/S8755-7223(03)00134-0).
- Williams, R. L., Oliver, R., & Stockdale, S. L. (2004). Psychological versus generic critical thinking as predictors and outcome measures in a large undergraduate human development course. *The Journal of General Education*, 53(1), 37–58. doi:[10.1353/jge.2004.0022](https://doi.org/10.1353/jge.2004.0022).

Zohar, A., & Nemet, F. (2002). Fostering students' knowledge and argumentation skills through dilemmas in human genetics. *Journal of Research in Science Teaching*, 39(1), 35–62. doi:[10.1002/tea.10008](https://doi.org/10.1002/tea.10008).

Dawit Tibebe Tiruneh is a doctoral student at the Faculty of Psychology and Educational Sciences of the KU Leuven, Belgium, and instructor at Bahir Dar University, Ethiopia. His main research interest is the design and development of learning environments for critical thinking.

Ataklti G. Weldelassie received his Ph.D. in Physics from KU Leuven and currently a teaching staff at the Science, Engineering and Technology Group, campus Group T of the KU Leuven, Belgium. His main research interest is in physics education.

Abrham Kassa is a lecturer in the department of Physics at Bahir Dar University, Ethiopia. His main research interest is in the teaching and learning of physics in higher education.

Zinaye Tefera is a lecturer in the department of Physics at Bahir Dar University, Ethiopia. His primary research interest is inquiry learning and the design of instructional strategies for promoting understanding in physics.

Mieke De Cock is associate professor in the department of Physics and Astronomy of the KU Leuven, Belgium, where she is responsible for the physics teacher training program. Her research focusses on conceptual understanding in physics, student use of mathematics in physics and integrated STEM education. She is teaching both introductory physics courses and teacher training courses.

Jan Elen is a professor of educational technology and teacher education at the Faculty of Psychology and Educational Sciences of the KU Leuven, Belgium. His main research interest is in the field of instructional design. He teaches both introductory and advanced courses in instructional psychology and educational technology. He is the senior editor of *Instructional Science*.